Vascular-targeted combination therapies for the treatment of cancer

Weiss, A.

2015

Chapter 5

Discovery of a low order drug-cell response surface for application in personalized medicine

Xianting Ding[1], Wenjia Liu[1], Andrea Weiss[2,3], Yiyang Li[1], Ieong Wong[4], Arjan W. Griffioen[3], Hubert van den Bergh[2], Hongquan Xu[5], Patrycja Nowak-Sliwinska[2] and Chih-Ming Ho[4].

[1] School of Biomedical Engineering, Med-X Research Institute, Shanghai Jiao Ton University (SJTU), Shanghai, China.

[2] Institute of Chemical Sciences and Engineering, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland.

[3] Angiogenesis Laboratory, Department of Medical Oncology, VU University Medical Center, Amsterdam, The Netherlands.

[4] Department of Mechanical and Aerospace Engineering, University of California, Los Angeles (UCLA), Los Angeles CA, USA.

[5] Department of Statistics, University of California, Los Angeles (UCLA), Los Angeles, CA, USA.

Phys. Biol., 2014, 11(6):065003.

Abstract

The cell is a complex system involving numerous components, which may often interact in a non-linear dynamic manner. Diseases at the cellular level are thus likely to involve multiple cellular constituents and pathways. As drugs, or drug combinations, may act synergistically on these multiple pathways, they might be more effective than their respective single target agents. Optimizing a drug mixture for a given disease in a specific patient is particularly challenging due to both the difficult selection of the drug mixture components to start out with and the all-important doses of these drugs to be applied. For N concentrations of M drugs, in principle, N^M combinations will have to be tested. As this may lead to a costly and time-consuming investigation for each individual patient, we have developed a Feedback System Control (FSC) technique which can rapidly select the optimal drug-dose combination from the often millions of possible combinations. By testing this FSC technique in a number of experimental systems representing different disease states, we found that the response of cells to multiple drugs is well described by a low order, rather smooth, drug-mixture-input/drug-effect-output multidimensional surface. The main consequences of this are that optimal drug combinations can be found in a surprisingly small number of tests, and that the translation from *in vitro* to *in vivo* is simplified. This points to the possibility of personalized optimal drug mixtures in the near future.This unexpectedly simple input-output relationship may also lead to a simple solution for handling the issue of human diversity in cancer therapeutics.

Introduction

The perturbation of cell homeostasis due to genetic and/or epigenetic changes can result in aberrant proteins or cell organelles. These can lead to uncontrolled cell growth, providing the underlying basis for most morbid and mortal illnesses. Modern drug discovery mainly aims to identify novel drug molecules which directly bind and inhibit such aberrant molecular cell targets[1,2]. One of the main challenges of drug discovery, therefore, is to identify such drug targets in a complex cellular system.

Biological systems, on a cellular, organ or body level, can be considered as complex systems[3]. Complex biological systems, however, are very different from other types of systems, such as an engineering system. A complex system consists of a large number of building blocks, e.g. proteins, mRNA, organelles. In biological systems, some of these functional molecules are self-organized into pathways[4,5]. In contrast, engineering systems are assembled with parts that are manufactured by following the first principle[6]. The system's functionality is well defined according to the design goal. On the other hand, cellular system level response emerge from a network of regulatory and signaling pathways and are adaptive with a large dynamic range[7]. Obviously, sorting out an aberrant cellular component for drug targeting in the midst of a complex system is like finding a needle in a haystack. Even if a target is identified and a drug is developed to inhibit it, single drug treatment often leads to drug resistance[8,9]. Furthermore, in many diseases it is common for there to be more than one disease causing target due to non-linear interactions between signaling pathways. Therefore, combinations of synergetic drugs targeted to several pathways and administered at low dose could represent an efficacious treatment strategy[10,11].

The efficacy of a drug combination not only depends on the selection of the drugs, but also on the dose ratios among the drugs[10]. M drugs with N dose levels will generate N^M possible combinations. A brute force search for an optimal drug combination in such a large parameter space is a prohibitive task. The recently developed Feedback System Control (FSC) technique[12] can direct biosystems toward a desired phenotypic outcome based on combinatorial drug stimulation. FSC can home in an optimal drug combination with several orders of magnitude less experimental efforts than testing all of the N^M possibilities. FSC takes a top-down systems approach by focusing on improving a phenotype based on varying the combinatorial input stimuli. This method completely avoids the bottom-up approach frequently used in biology, where one attempts to predict and control cell behavior based on an understanding how different signaling pathways and molecules interact. It is surprising that typically less than 15 iteration loop tests can

identify the optimal combination from millions or more alternatives. FSC is a platform technology which has been demonstrated in the treatment of cancers[13], inhibition of viral infection[10], the maintenance of human embryonic stem cells (hESC)[11], the reformulation of Chinese herbal medicine[14], and the differentiation of mesenchymal stem cells[15].

Even though it took a long time and a lot of effort, many new targeted drugs and their combinations have been introduced to the clinics in the past three decades. Unfortunately, patients' response rates to most targeted drugs remain fairly low in cancer treatments. For example, the response rate for lung cancer patients is about 25% and only 10% for hepatoma[16]. Many reasons contribute to these unsatisfactory results. Patient diversity and cancer heterogeneity are among some of the factors which influence the efficacy cancer therapy[17]. Genetic profiles of individual patients with the same disease vary across genders, races, etc. and causes diversities of proteomic network through transduction. The current clinical practice for chemotherapy is to use the same regimen for the patients with the same type of disease[18,19] , therefore, relatively low response rates are frequently observed.

With the rapid development of micro/nano technology based diagnostic instrument, fast and affordable genetic analyses have become available and this has enabled the development of genotypic personalized medicine (GPM)[20,21]. GPM is based on the principle of customizing single targeted or combinatorial drugs for a group of patients with similar gene profiles and can result in better therapeutic outcomes. These strategies, however, disregard the fact that disease can also be independently caused by epigenetic stimulations[22,23]. Therapeutic procedures can obviously be much more precise, if they include the consideration of disease phenotypes. However, phenotypic personalized medicine (PPM) needs to have quantitative efficacy-drug relationship *a priori*.

In this paper, we will present the results from investigations on four biological models, including non-small-cell lung cancer (NSCLC) treatment, Herpes Simplex Virus type 1 (HSV-1) eradication, mesenchymal stem cell osteogenesis induction and cancer angiogenesis inhibition. In these studies, we show that the efficacy-drug dose relationships of each system are simple and smooth. This finding comes from one of the fundamental characteristics of complex system. Due to the process of evolution, organisms have developed in such a way that they are robust and adaptive to environmental stimulations. That is, the bio complex system response surface to extracellular stimulations must be very smooth. This is the reason why we can easily locate an optimal drug combination in a biological system after approximately 15 iterations of the FSC

technique. By testing a small group of subjects, the efficacy-drug surface can be established. With this quantitatively defined relationship, PPM can be practiced with great confidence.

Results

Non-Small-Cell Lung Cancer (NSCLC) reduction

The authors of this study[13] aimed to optimize the combination of three anti-cancer drugs: AG490, a Janus Kinase 2 inhibitor; U0126, a MEK1 and MEK2 inhibitor; and Indirubin-3'-monoxime, which is an antimitotic CDK/GSU inhibitor. Compounds like the latter are sometimes used as an ingredient in Chinese medicine. As a measurement of the efficacy of the combination treatment and its 'selectivity', the ATP levels of both an NSCLC cell line (A549) and of a normal primary lung fibroblast (AG02603), are measured after exposure to a large number of combinations of the three above-mentioned agents for 72 hrs. It is argued that the three selected drugs target distinct but nevertheless somewhat connected intracellular pathways related to cell survival and proliferation. Eight doses were selected for each of the three drugs resulting in 512 possible combinations. The authors test all these combinations and report the results in this study.

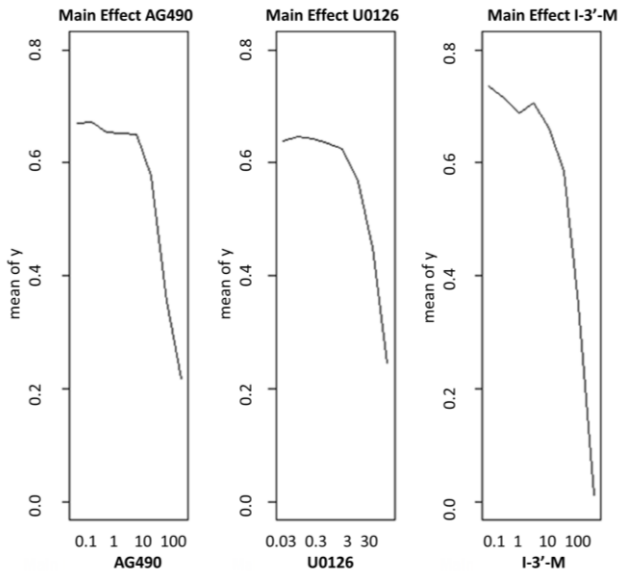


Figure 1. Single drug effects for the three drugs used in the Non-Small-Cell Lung Cancer (NSCLC) reduction study. The plots show the change in system readout 'y', representing the percentage of surviving cancer cell, as a function of the concentration (indicated on the x-axis in μM) of each drug.

We then further analyzed the 512 data points by building a second order linear regression model. At first we looked at the impact of the dose change of each individual drug on the therapeutic output 'y' which quantifies cancer cell survival. These single drug effects, also known as 'main effects' in statistics are shown in Fig. 1. All three drugs demonstrate significant main effects, as can be seen by the obvious decrease of 'y' with increasing dose of each compound.

We then checked the fidelity of the linear regression model built from the 512 data points and found $R^2 = 0.96$ which is indicative of the good correlation between the regression model and the experimental observations. We then plotted the model predictions and the experimental points side by side in Fig. 2a. The fact that all points fall on a line again suggests that the regression model faithfully represents the experimental data. In Fig. 2b we then generated a residual plot to evaluate if the model might be biased for any particular fitted values. The residuals are all fairly close to zero, indicating that the regression model used is not biased. We then generated a 'normal Q-Q plot' to see if the residual data follow a normal Gaussian distribution. This is shown in Fig. 2c. If the residual data do follow a normal distribution, the points on the Q-Q plot will fall approximately on a straight line. This normal Q-Q analysis is another way of checking if the regression model used shows bias. If the residual plot does not follow a normal distribution then we will have to re-examine the data using another modeling strategy. In this case however, the residual points fall close to a straight line, suggesting that the assumption of a normal distribution holds. Only a few points fall far from the straight line. These few points are unlikely to significantly influence the overall statistical analysis and can be treated as statistical outliers. In order to tell which data points are likely to be outliers (e.g. points that are due to errors in the experimental measurements), we generated the Cook's distance for each data point as shown in Fig. 2d. Cook's distance is a well-accepted way to tell whether any outliers exist in the data set. Data point No. 337 clearly shows a Cook's distance value larger than any other data point. This indicates that this particular point should be analyzed carefully and possibly repeated, as it is likely that a measurement or recording error occurred.

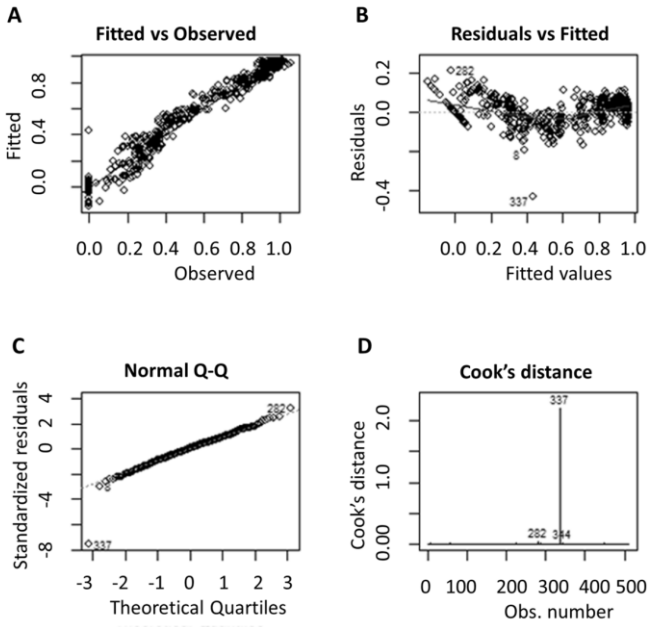


Figure 2. Analysis of the regression model generated using the combinatorial data produced in the NSCLC reduction study. (A) The model's predicted values plotted against experimentally observed data points; (B) the residual plot shows a mean (the red line) close to zero, indicating that the model is not biased at any particular drug doses; (C) the Normal Q-Q plot; and (D) the Cook's distance plot.

One key characteristic of linear regression modeling, especially lower order regression models is that the response surfaces can be depicted with smooth contours, at least in the region of the experimental values reported. In Fig. 3, contour plots of each of the two-drug pairs out of the three drugs tested are generated while fixing the third drug's dose to be equal to zero. Thus when I-3'-M is not added to the mixture, AG490 and U0126 combined optimally when AG490's dose is between 200 and 250 μM, while U0126's dose is between 80 and 100 μM, possibly indicating some degree of synergism. On the other hand, when the concentration U0126 is set to zero, and AG490 and I-3'-M are applied together, the optimal anti-cancer effect is obtained when both drugs are at the high dose end, i.e. no synergy is observed between these two drugs. Finally, when the concentration of AG490 is set to zero, and we look at U0126 and I-3'-M being used together, the optimal anti-cancer effect occurs when the U0126 concentration is between 100 and 150 μM, and either higher or lower doses of this compound lead to a less optimal readout. The three contour plots generated all appear to be quite smooth in the domain of concentrations tested, confirming that the relation between the three chosen anti-cancer drugs can be expressed with a simple second order linear regression model.

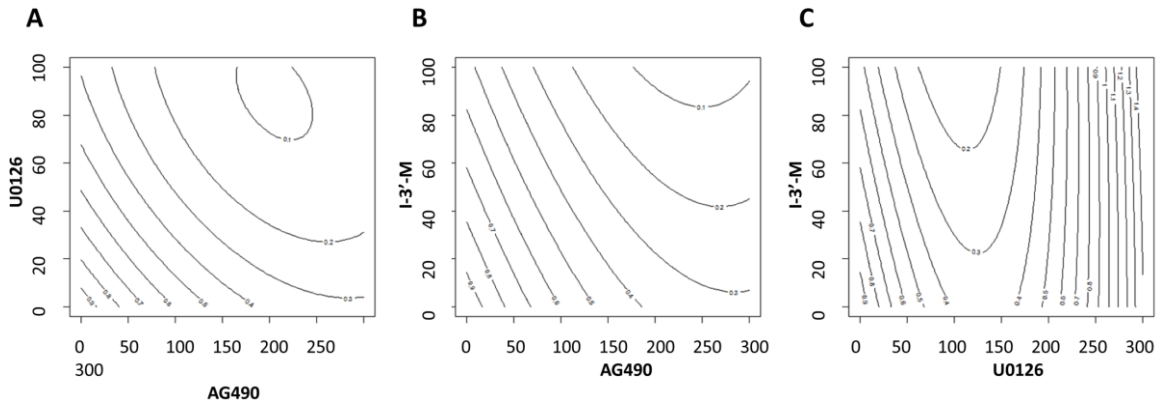


Figure 3. Smooth contour plots of any two pairs of the three drugs used in the NSCLC reduction study. (A) AG490 vs. U0126; (B) AG490 vs. I-3′-M; (C) U0126 vs. I-3′-M. The contour plots are continuous without any sudden breaks, confirming the smoothness of the drug combination dose-efficacy relationship.

Herpes Simplex Virus type 1 (HSV-1) infection

In the first example above, we examined a typical anti-cancer system treated with combinations of three drugs. In this second example we investigate viral infection, another kind of complex biological system which is commonly treated with a drug mixture. In these viral-based diseases the drugs can interact with both the host cells and the virus itself, thus increasing the complexity of the system. We were now interested to see if in such a complex biological system the relation

between the drug doses and the 'y' readout, which is the percent of viral infection, can still be described by a 2nd order linear regression model. We thus analyzed the measured percentage of viral infection after treatment with different drug combinations in a Herpes Simplex Virus type 1 (HSV-1) infected system[10]. In this paper, Ding et al. aimed to optimize combinations of the following drugs at seven drug dose levels: Interferon-alpha, Interferon-beta, Interferon-gamma, Ribavirin (a guanosine analog which interferes with viral RNA synthesis), Acyclovir (a drug which is converted in the body to a strong inhibitor of viral DNA-polymerase), and TNF-alpha. The drugs were tested as to their efficacy in eradicating an HSV-1 infection in NIH 3T3 fibroblasts. Six drugs applied at seven dose levels leads to a total of 117,649 possible combinations. With the help of the FSC technique described above, the authors attempted to identify the most effective drug combinations for inhibiting the viral infection. In this case only 192 drug combinations were tested through 12 rounds of experimental measurements, and the FSC method permitted us to find several drug combinations that effectively and completely eradicate the signs of viral infection. Due to the fact that Ribavirin has a large number of possible side effects which include bone pain, increased stomach acid and blurred vision, a second FSC search was undertaken without Ribavirin. The latter permitted us to identify an effective Ribavirin-free drug combination in only 20 rounds of otherwise the same experimental tests.

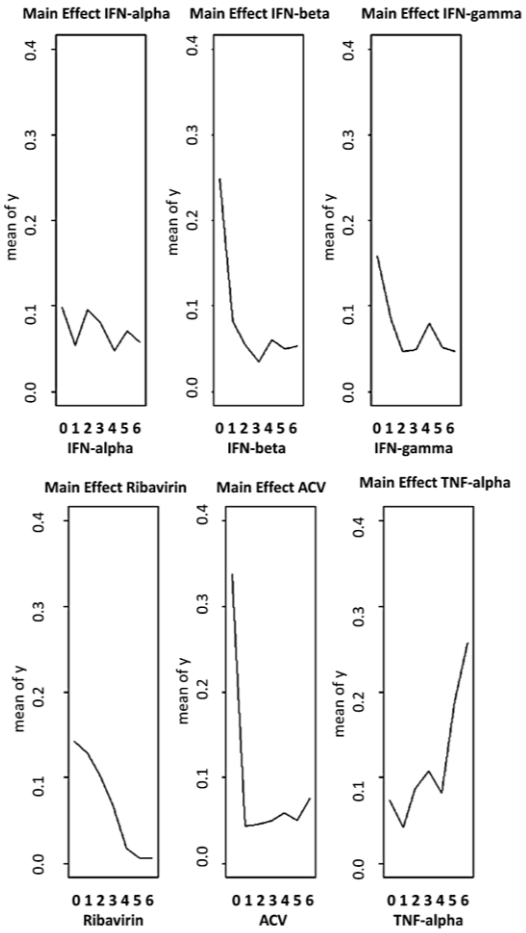


Figure 4. Single drug effects for the six drugs used in the Herpes Simplex Virus type 1 (HSV-1) infection study. Plots show the variation of the system readout 'y', representing the percentage of virally infected cells, as a function of varying the concentration of a given drug. The concentrations of each drug are indicated on the x-axis and are represented as coded concentration levels 1-6.

In the current study, we further analyzed the drug combination data achieved from the previous FSC search. To do this we investigated how the change of the dose of each individual drug would influence the percent viral infection, i.e. the influence of one particular selected drug on the overall system output 'y' (percentage of viral infection). In such 'main effect plots', obtained from the pooled output data, we fix the dosages of the other drugs and vary only the dose of the drug of interest. In most cases, an increase in drug dose led to a better output, meaning a lower value of the percentage of viral infected cells. The only exception was TNF-alpha, which shows the opposite effect, i.e. where a dose increase led to a higher percentage of viral infection (Fig. 4).

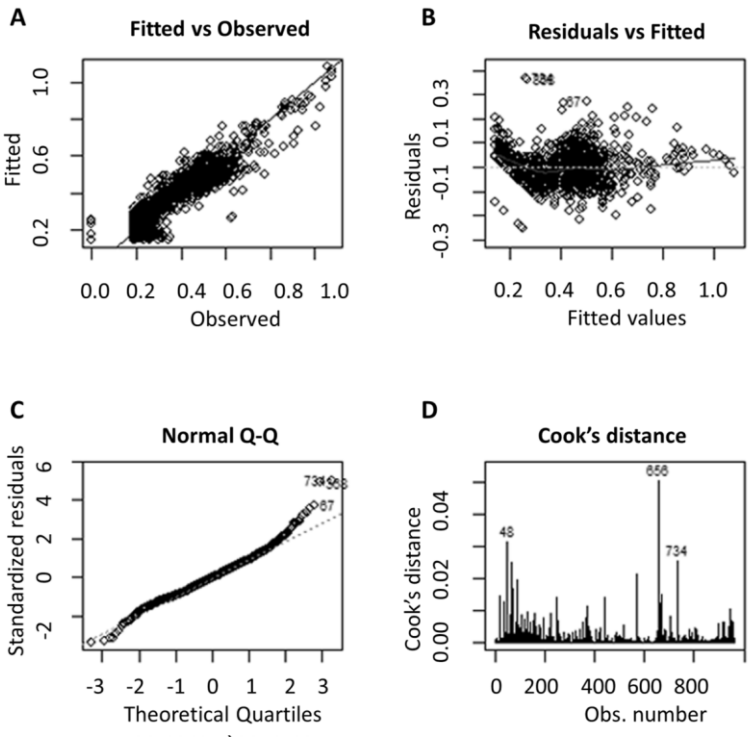


Figure 5. Regression analysis of the model generated from the HSV-1 infection study data. As the system readout 'y', representing the percentage of virally infected cells, was non-normally distributed, a 'Boxcox' transformation on 'y' was made to transform the 'y' data set to an approximately normal distribution before the regression analysis was done. (A) Plot of the model fitted values versus the experimentally observed values; (B) the residual plot; (C) the Normal Q-Q plot; (D) the Cook's distance plot.

We then generated a 2^{nd} order linear regression model for 'y' with the six drugs as independent variables. The model yielded a value for R^2 = 0.7448. The fidelity is not as good as what was seen in the NSCLC case. However, for a biolgoical system as complex as viral infection, where the internal variances, such as experimental batch-to-batch variance can be as large as 20%, this is not unreasonable. We then performed a 'boxcox' transformation on 'y' data. Boxcox transforms non-normally distributed data to a set of data that has approximately normal distribution. After the boxcox transformation, a fourth root transformation was made on the 'y', and the regression analysis was repeated by regressing $y^{1/4}$ over the six drugs. This transformation resulted in an increase of R^2 from 0.7448 to 0.8103. In Fig. 5a, we show that the 'model-fitted' values and the experimental values for 'y'

agree quite well with one another. As in the case of NSCLC above, we also examined the residual plot and the Normal Q-Q plot for this model. These data are shown in Fig. 5b and 5c, and showed no clear indications suggesting failure or bias of the model. The Cook's distance plot suggested only a few outliers, as can be seen in Fig. 5d. Therefore one may conclude that even in such a complex case as viral infection, the effect of a drug mixture on the percentage of viral infected cells can still be modeled with a 2nd order linear regression, though with somewhat reduced confidence.

Mesenchymal Stem cell osteogenesis induction

As a third example of a biologically complex system to optimize drug mixtures for, we chose to investigate a case of chemical stem cell osteogenesis induction. Chondrogenic differentiation of mesynchymal stem cells from bone marrow into mature tissue cells has been shown to be mainly sensitive to the intrinsic properties of the extracellular matrix like its structure, elasticity and composition. Yoshitomo et al. however decided to study the application of combinations of seven chemical compounds which are, among others, known to promote the induction of osteogenic differentiation of mesenchymal stems cells (MSC)[15]. It might thus be expected that the relation between the dose of these extrinsic chemicals and the system readout 'y', which in this case is osteogenic cell differentiation, might not follow a smooth relation as given by a simple 2nd order linear regression. The following seven compounds were included in the investigation: AA2P (L-

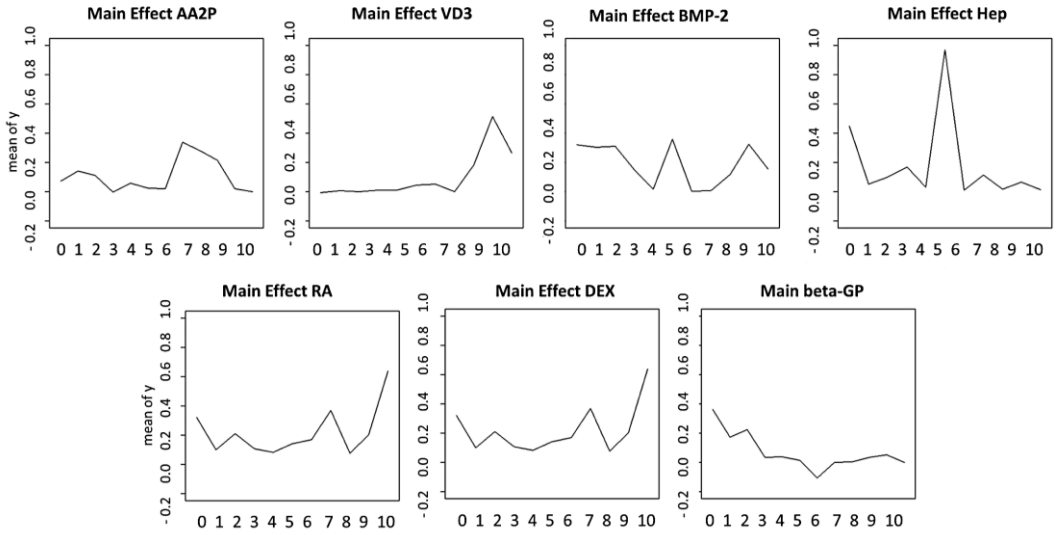


Figure 6. Single drug effects for the seven drugs used in the mesenchymal stem cell osteogenesis induction study. Plots show the variation of the system readout 'y', representing the percentage of cells to undergo osteogenic differentiation, as a function of varying the concentration of a given drug. The concentrations of each drug are indicated on the x-axis and are represented as coded concentration levels 1-10.

ascorbic acid 2-phosphate), VD3 (Vitamin D3), BMP-2 (bone morphogenic protein 2), RA (retinoic acid), Dex (dexamethasone) and beta-GP (beta-glycerophosphate). The authors of this study applied the FSC technique and tested 107 drug combinations experimentally. This led to a unique combination of drugs that robustly induces bone mineralization.

In the present study, we further analyzed the same data set as was generated by the FSC search by applying second order linear regression analysis. As in the previous cases, we first looked at the influence of each individual drug on the output 'y'. This is shown in Fig. 6. It is interesting to note that in this particular example of a complex biological system, when the concentrations of the other chemicals are fixed, the influence of the individual concentrations of each single compound on 'y' appears to be quite diverse. No simple trend can be observed for any of these individual compounds. This observation led us to believe that the optimal chemical mixture for osteogenesis would be a rather unique combination of compounds, in other words that the relation between 'y' and the doses of the compounds would not follow a well-defined 2nd order linear regression model (i.e. the surface would not be smooth but rather be characterized by a quite localized extreme value).

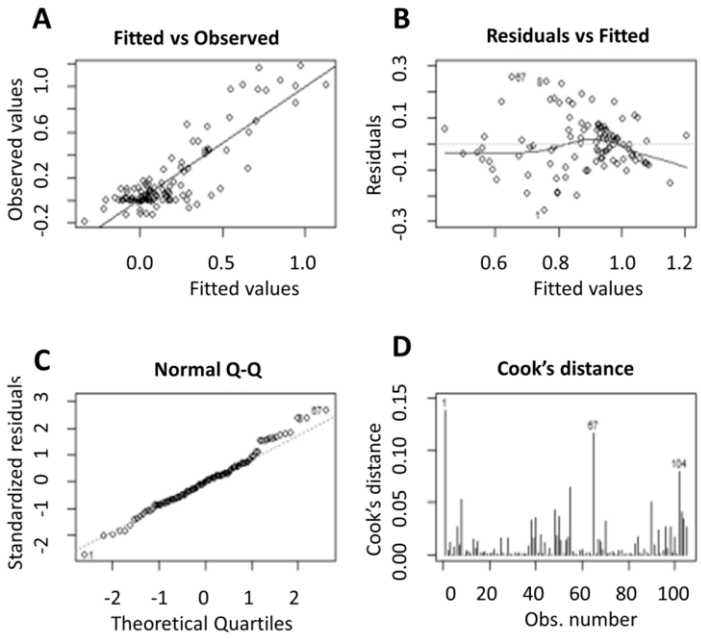


Figure 7. Analysis of the regression model generated based on the data obtained in the Mesenchymal Stem cell osteogenesis induction study. (A) The model's predicted values plotted against experimentally observed data points; (B) the residual plot; (C) the Normal Q-Q plot; and (D) the Cook's distance plot.

Yet, to our surprise, the 2nd order linear regression model generated based on the tested data points showed an R^2 value of 0.7444, which was much higher than our expectation. When we examined the residual plots and Normal Q-Q plots, we observed a pattern in residual plots, which indicated the residuals did not obtain a mean of zero (Supplementary Fig. S1b). When we further looked at the Cook's distance plot (Supplementary Fig. S1d), we found data point No. 45 was a clear outlier.

We then repeated the whole analysis by removing data point No. 45. Indeed, the experimental observations and the model predictions basically agree with one another, as can be seen from Fig. 7a. Furthermore, upon examination of the residual plots and the Normal Q-Q plots (shown in Fig. 7b and 7c), we did not observe a clear residual distribution, indicating that even though there was only a modest fidelity, the second order linear regression model was an acceptable approximation to the experimental data. Fig. 7d shows that Cook's distance analysis did not indicate any new outliers either.

Cancer angiogenesis inhibition

After examining the input-output relation for the efficacy of multiple drug treatments *in vitro* in the three different complex biological systems described above, we found that the system's response can in each case be fairly well modeled with a smooth second order equation. If this finding would be universally true for other complex biological systems, it may be possible to identify optimal drug combinations by testing a limited number of data points, which are then fitted with a second order response function. This function can then be used to identify the optimal drug combination.

In order to test this rather bold hypothesis, we tried to find an optimal combination of a fairly small number of anti-angiogenic drugs (if possible 4 or less), starting from 9 such substances. The starting compounds were carefully selected for their known or presumed complementary anti-angiogenic activity. The drugs chosen were: Anginex (1), Avastin (2), Axitinib (3), Erlotinib (4), Anti-HMGB1 antibody (5), Sunitinib (6), Anti-Vimentin antibody (7), RAPTA-C (8), and BEZ-235 (9). As, up to this date, such treatments have been only partially successful with the single components listed above, it is now hoped that by using the FSC technique an effective angiostatic combination of some of the above listed drugs might be found. If, furthermore, this combination would benefit from a significant degree of synergism, it cannot be excluded that the treatment would also imply the use of much lower drug doses than are normally applied with the single compounds. Reduced drug doses may also carry the potential for having fewer side effects and reduced driving force for the development of drug resistant. Four drug concentrations were assigned to each of the 9 starting compounds thus creating $4^9 = 262144$ possible drug combinations (Table 1).

Immortalized human macrovascular endothelial cells (ECRF24) were used in this study. We first applied the FSC technique to optimize these nine drugs over ten iterations. The FSC search optimization will be published in a separate paper as the present article focuses mainly on the

analysis of the cell survival response to the different drug doses. In each of the 10 iteration steps, 19 drug combinations were tested, giving a total of 190 data points. Like before we used these points to build a second order linear regression model. As expected, the predictions from the regression model and the data points showed a high correlation of 0.8642. The R^2 value equals 0.7468.

Based on the regression model shown in equation (1) below (see Method session), where the output 'y' is a function of k drugs, we then tried to find the optimal drug combination by looking carefully at all the single drug coefficients (β_i), all the drug-drug interaction coefficients (β_{ij}), and all the single drug quadratic regression coefficients (β_{ii}) which are plotted in Fig. 8a. Since the model and the experimental data showed a very good correlation, we then used Fig. 8a to help us eliminate less important drugs. We also take into account that, as shown in Fig. 8b, the Normal Q-Q plot shows there seems to be no violation of the normality hypothesis for the residuals distribution. The y-axis of Fig. 8a is such that large negative values of coefficients plotted along the y-axis imply a better

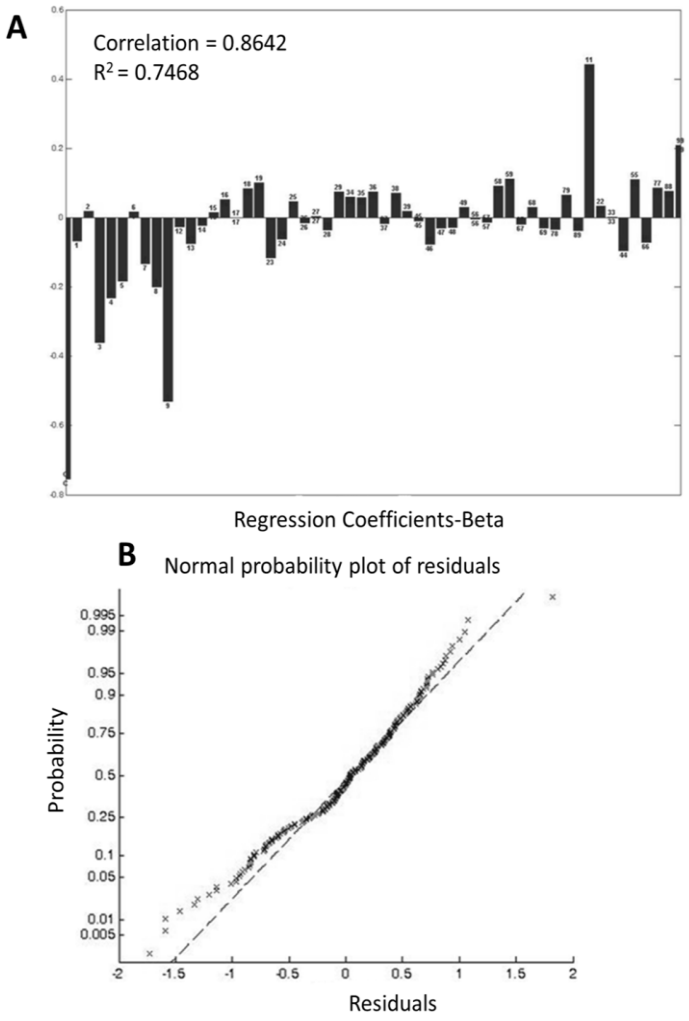


Figure 8. Regression analysis for the study of cancer angiogenesis inhibition. (A) All of the regression coefficients obtained from the regression model of the angiogenesis inhibition study are presented. Regression coefficients were used to eliminate the less significant drugs used in the final combination; (B) the Normal Q-Q plot is provided to verify residual normality.

contribution to lowering endothelial cell viability, and thus a higher angiostatic effect. Anginex (1) showed a relatively low single drug effect (β_1) but a much higher 2nd order effect (β_{11}), meaning Anginex (1) is not likely to be a highly effective drug when used individually. Furthermore, Anginex (1) did not show the desired very low values in Fig. 8a for the interaction with the other drugs. This observation may be interpreted as minimal 'synergies' between Anginex and the other drugs (i.e. large negative values of β_{1k}). Based on these reasons, we could confidently drop Anginex (1) from the drug mixture. Moreover, Avastin (2) and Sunitinib (6) did have slightly positive coefficients (β_2 and β_6) in Fig. 8a (although this may not be statistically very significant), which is indicative of not very good single drug contributions to the angiostatic effect. As the model is run with real concentrations of each drug, single drug effects are quite important. We therefore dropped drugs (2) and (6). The Anti-HMGB1 Antibody (5) has a quite negative coefficient for its single drug effect (β_5). However, unfortunately this drug showed mainly 'antagonistic-like' effects (the β_{5k} coefficients are largely positive or around zero in Fig. 8a with other drugs), and it had a rather positive value quadratic term (β_{55}). Therefore, drug (5) was also dropped from the drug combination. Finally, the Anti-vimentin Antibody (7) also shows a good single drug effect, but a fairly positive 2nd order effect, as well as mainly small coefficients for the interactions with other drugs, indicating slight 'synergism' or slight 'antagonism'. We therefore also dropped drug (7) at this point. These procedures lead us to the optimal combination including the following four drugs: Axitinib (3), Erlotinib (4), RAPTA-C (8) and BEZ-235 (9). In later analysis, axitinib was further eliminated from this mixture, for various reasons including toxicity.

Thus the final optimal drug combination consisted of drug (4) (1uM) + (8) (100uM) + (9) (0.02uM). This combination, at these rather low concentrations was sufficiently potent to inhibit more than 90% of the endothelial cell proliferation. Note that the EC50 values (the values that gave 50% survival) of drug (4), (8), (9) are approximately 20 μM, 500 μM and 0.05 μM, respectively, while the drug doses used in combination were much lower. This fact confirms the successful application of the 2nd order linear regression modeling in selecting a small group of very well-interacting angiostatic compounds from the original nine. It also points to the possibility of rapidly designing effective drug combinations, an opportunity that will be further tested *in vivo* and in preclinical models in the near future. One may speculate that such a rapid and not excessively expensive method of individual drug optimization may be quite useful in the case of cancers which change rapidly as a function of time.

Anginex	Avastin	Axitinib	Erlotinib	Anti-HMGB1 Ab	Sunitinib	Anti-Vimentin Ab	RAPTA-C	BEZ-235
0	0	0	0	0	0	0	0	0
0.13	0.007	0.01	0.1	0.02	0.05	0.09	0.05	0.0005
0.76	0.07	0.3	0.5	0.09	0.1	0.17	1	0.001
1.8	0.1	1	2	0.17	0.5	0.26	5	0.005

Table 1. Single drug concentration table used in anti-angiogenesis study (units in μM).

Discussion

The diseased state in a cell commonly involves a number of abnormal signaling pathways. It is therefore unlikely that a single drug could inhibit all of the aberrantly activated pathways involved in a disease and its progression. Although in some cases, individual drugs can show satisfying efficacy in the treatment of a disease, the toxicity induced when drugs are used at high dose, and drug resistance accumulated from long-term drug administration, still often limit single-drug regimens as an effective long-term treatment option. An alternative approach is to use drug combinations that can simultaneously target multiple diseased cellular nodes. The synergy among drugs in an effective combination can lead to reduced dose, relieved side-effects and increased efficacy. In all the cases we studied, optimized drug combinations were superior to their single-drug counterparts.

The FSC approach usually identifies the optimal drug combination in less than 15 rounds of experimental efforts, by testing only 2% or less of the total possible search space. This paper, by investigating four different biological systems, demonstrates that although biological systems are internally complex, the drug dose-efficacy relationships can frequently be expressed by low order input-output multi-dimensional surfaces. This finding not only explains the puzzle of why FSC is effective in drug combination optimization, but also serves as the foundation for the idea that a small number of well-designed experimental tests are adequate to form a reasonable response surface for predicting optimal drug combinations and doses.

In a biological system with multiple factors, experimentally testing all the possible drug combinations can be a very laborious, time consuming and costly process. If the response of a biological system can be described with a smooth function, then we only need to perform a small number of tests in order to build up a reliable model for this input-output relationship. This will

then allow the rapid examination of the entire search space. This suggests that to optimize combinations of multiple factors in a bio-complex system, we may start with only a few tests and examine whether the bio-system's response is smooth or not. If so, the optimal combination could be faithfully designed by building up the smooth response surface of the system with relatively few test data points.

Clinically, tumor progression at different time points and at different location may require different therapies for effective treatment. Even within the same tumor, multiple subtypes of cells could require distinct treatments. This fact emphasizes the necessity of applying combinatorial drugs to treat lethal disease. Tumor, as well as the other complicated diseases, often involve multiple subtypes/strains or pathogenic intracellular signaling pathways. This complexity varies with time and location as well. It is extremely challenging for one single compound to deal with such complexity. Drug combinations often tackle the problem from different angles, and therefore, are believed to be a more universal and effective solution. The aim of this paper is to reveal the fact that although biological systems are internally complex, the relationships between drug doses and phenotypic system responses often follow relatively simple patterns. Furthermore, these patterns can often be modeled faithfully using second order regression analysis. In order to verify this conclusion, the authors have tested the hypothesis in four different biological systems. These biological systems were selected 'on purpose' to be different in order to cover a relatively large interest of biological research.

Genotypic personalized medicine (GPM) has been extensively discussed recently. GPM diagnoses and categorizes patients based on their genotypic traits, and then treats them based on targeted strategies. GPM has greatly improved response rates to treatment in cancer patients. Yet, epigenetic stimulation can also independently lead to disease. Therefore, genetic investigations only address part of the problem underlying diseased states. For this reason, developing therapeutic strategies based on phenotypic clues may in fact be a more direct route to evaluating the efficacy of a treatment. However, in order to practice phenotypic personalized medicine (PPM), quantitative efficacy-drug relationships need to be understood *a priori*. The work presented here demonstrates the fact that a low order drug-cell response surface may well commonly exist in many biological systems, indicating that the FSC approach could be an invaluable route towards PPM optimization.

Acknowledgement

This work was supported by National Natural Science Foundation of China (81301293) and National Science and Technology Major Projects for "Major New Drugs Innovation and Development" (2014ZX09507008).

Materials

Regression modeling

Regression modeling is done with R© and MATLAB© programming languages.

For a bio-complex system with k drugs, a standard form of the linear regression model is as follows:

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} {x_i}^2 + \sum_{i=1}^{k} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j + \varepsilon \qquad (1)$$

where β_0, β_i, β_{ii} and β_{ij} are the intercept, linear, quadratic and bilinear (or interaction) terms [24,25]. In this study, a full model with all the coefficients (including intercept, linear, quadratic and interaction terms) in a 2nd order linear regression model was built for each case.

A stepwise linear regression analysis was done in R programming to remove those statistically non-significant regression terms to form a cleaner final regression model. Fitted values and experimentally observed values are plotted side-by-side to evaluate the fitness of the regression model. Residual plot was generated to evaluate whether the regression model is biased for any particular fitted values. Cook Distance plot was generated to tell possible outliers in the experiment, if any. Finally, a series of transformation (log/square-root/square transformation) on the system readout was made to increase the fitting efficiency (P-value) of the regression model.

Data sets

Three datasets were selected in this study based on previously published literature [10,13,15]. In all these three cases, the FSC technique was applied to identify optimized drug cocktails to tackle different biological questions. This paper aimed to study why the FSC technique could be implemented so effectively to optimize drug combinations, so we selected to only analyze data sets that were generated using the FSC drug cocktail search practice. The data set in the anti-angiogenesis study was generated by the authors in a separate experiment to validate the findings from the study of the first three data sets.

Drugs acquisition for anti-angiogenesis study

Anginex® was provided by Peptx (Excelsior, MN, USA). Erlotinib and Axitinib were obtained from LC laboratories (Woburn, MA, USA), Sunitinib was from Pfizer Inc. (New York, NY, USA) and BEZ235 was from Chemdea LLC (Ridgewood, USA). Anti-vimentin monoclonal mouse antibody (clone V9) was purchased from Dako (Glostrup, Denmark) and anti-HMG1 antibody was purchased from Santa Cruz Biotechnology (Heidelberg, Germany). Avastin® was purchsaed from Genentech (San Francisco, CA, USA). RAPTA-C was synthesized and purified based on previous publication [26].

Cell culture and maintenance for anti-angiogenesis study

Immortalized human vascular endothelial cells (ECRF24) were cultured in cell culture medium containing 50% DMEM and 50% RPMI 1640 supplemented with addition of 1% antibiotics (Life Technologies, Carlsbad, California, USA).

Cell viability assay for anti-angiogenesis study

Cells were seeded at density of 2.5 x 10^3 cells/well on a 96-well culture plate. Cells were allowed for 72 h incubation time with drug combinations. Drugs were premixed in culture medium. Cell viability was calculated using the CellTiter-Glo luminescent cell viability assay (Promega, Madison, WI, USA).

References

1 Imming, P., Sinning, C. & Meyer, A. Drugs, their targets and the nature and number of drug targets (vol 5, pg 821, 2006). *Nat. Rev. Drug Discov.* **6**, 126-126 (2007).

2 Zheng, C. J. *et al.* Therapeutic targets: Progress of their exploration and investigation of their characteristics (vol 58, pg 259, 2006). *Pharmacol. Rev.* **58**, 682-682 (2006).

3 Ambrosini, E. & Aloisi, F. Chemokines and glial cells: A complex network in the central nervous system. *Neurochem. Res.* **29**, 1017-1038 (2004).

4 Steinman, R. M. The Dendritic Cell System And Its Role In Immunogenicity. *Annu. Rev. Immunol.* **9**, 271-296 (1991).

5 Fields, S. & Song, O. K. A Novel Genetic System To Detect Protein Protein Interactions. *Nature* **340**, 245-246 (1989).

6 Steward, D. V. The Design Structure-System - a Method for Managing the Design Of Complex-Systems. *IEEE T Eng. Manage.* **28**, 71-74 (1981).

7 Alon, U. Biological networks: The tinkerer as an engineer. *Science* **301**, 1866-1867 (2003).

8 Fitzgerald, J. B., Schoeberl, B., Nielsen, U. B. & Sorger, P. K. Systems biology and combination therapy in the quest for clinical efficacy. *Nat. Chem. Biol.* **2**, 458-466 (2006).

9 De Clercq, E. Antiviral drugs in current clinical use. *J. Clin. Virol.* **30**, 115-133 (2004).

10 Ding, X. T. *et al.* Cascade search for HSV-1 combinatorial drugs with high antiviral efficacy and low toxicity. *Int. J. Nanomed.* **7**, 2281-2292 (2012).
11 Tsutsui, H. *et al.* An optimized small molecule inhibitor cocktail supports long-term maintenance of human embryonic stem cells. *Nat. Commun.* **2** (2011).
12 Wong, P. K. *et al.* Closed-loop control of cellular functions using combinatory drugs guided by a stochastic search algorithm. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5105-5110 (2008).
13 Al-Shyoukh, I. *et al.* Systematic quantitative characterization of cellular responses induced by multiple signals. *BMC Syst. Biol.* **5** (2011).
14 Yu, H. *et al.* Optimizing Combinations of Flavonoids Deriving from Astragali Radix in Activating the Regulatory Element of Erythropoietin by a Feedback System Control Scheme. *Evid-Based Compl. Alt.* (2013).
15 Honda, Y. *et al.* Guiding the osteogenic fate of mouse and human mesenchymal stem cells through feedback system control. *Sci. Rep.-U.K.* **3** (2013).
16 Depierre, A. *et al.* A Phase-Ii Study Of Navelbine (Vinorelbine) In the Treatment Of Non Small-Cell Lung-Cancer. *Am. J. Clin. Oncol-Canc.* **14**, 115-119 (1991).
17 Shackleton, M., Quintana, E., Fearon, E. R. & Morrison, S. J. Heterogeneity in Cancer: Cancer Stem Cells versus Clonal Evolution. *Cell* **138**, 822-829 (2009).
18 Ford, D. *et al.* Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. *Am. J. Hum. Genet.* **62**, 676-689 (1998).
19 Klein, C. A. *et al.* Genetic heterogeneity of single disseminated tumour cells in minimal residual cancer. *Lancet* **360**, 683-689 (2002).
20 Wheeler, D. A. *et al.* The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**, 872-U875 (2008).
21 Hood, L., Heath, J. R., Phelps, M. E. & Lin, B. Y. Systems biology and new technologies enable predictive and preventative medicine. *Science* **306**, 640-643 (2004).
22 Zhang, H. Y. *et al.* Long-term expansion of human neural progenitor cells by epigenetic stimulation in vitro. *Neurosci. Res.* **51**, 157-165 (2005).
23 Vescovi, A. L. *et al.* Isolation and cloning of multipotential stem cells from the embryonic human CNS and establishment of transplantable human neural stem cell lines by epigenetic stimulation. *Exp. Neurol.* **156**, 71-83 (1999).
24 Ding, X. T., Xu, H. Q., Hopper, C., Yang, J. & Ho, C. M. Use of Fractional Factorial Designs in Antiviral Drug Studies. *Qual. Reliab. Eng. Int*. **29**, 299-304 (2013).
25 Jaynes, J., Ding, X. T., Xu, H. Q., Wong, W. K. & Ho, C. M. Application of fractional factorial designs to study drug combinations. *Stat. Med.* **32**, 307-318 (2013).
26 Allardyce, C. S., Dyson, P. J., Ellis, D. J. & Heath, S. L. [Ru(eta(6)-p-cymene)Cl-2(pta)] (pta=1,3,5-triaza-7-phosphatricyclo[3.3.1.1]decane): a water soluble compound that exhibits pH dependent DNA binding providing selectivity for diseased cells. *Chem. Commun.*, 1396-1397 (2001).

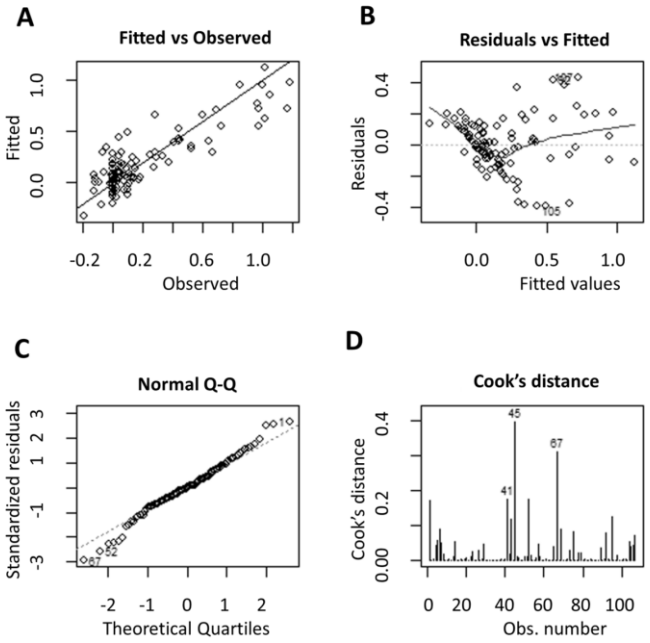
Supplementary Material

Discovery of a low order drug-cell response surface for applications in personalized medicine

Contents:

Supplementary figure.

Supplementary Figure



Supplementary Figure S1. Analysis of the regression model generated based on the data obtained in the Mesenchymal Stem cell osteogenesis induction study before the removal of outlier point 45. (A) The model's predicted values plotted against experimentally observed data points; (B) the residual plot; (C) the Normal Q-Q plot; and (D) the Cook's distance plot.